HUMAN FACTORS, 1986, 28(4), 421–438

Statistical Dependency in Visual Scanning

STEPHEN R. ELLIS,[1] *NASA–Ames Research Center, Moffett Field, California, and Department of Physiological Optics, University of California, Berkeley, and* LAWRENCE STARK, *Department of Physiological Optics, University of California, Berkeley*

A method to identify statistical dependencies in the positions of eye fixations is developed and applied to eye movement data from subjects who viewed dynamic displays of air traffic and judged future relative position of aircraft. Analysis of approximately 23 000 fixations on points of interest on the display identified statistical dependencies in scanning that were independent of the physical placement of the points of interest. Identification of these dependencies is inconsistent with random-sampling-based theories used to model visual search and information seeking.

INTRODUCTION

The distribution of eye fixations on stimuli in the visual field is usually not uniform, and the times spent viewing each of their component features are usually not equal (Buswell, 1935; Fisher, Monty and Senders, 1981; Fitts, Jones, and Milton, 1950; Papin, Naureils, and Santucci, 1980; Senders, Fisher, and Monty, 1978; Stark and Ellis, 1981; Yarbus, 1967). Certain features are often more "popular" than others. The resulting distribution of fixations may be expressed as a zero-order probability vector where each element is the probability of viewing a particular feature. When the scanning among the features is otherwise random, this vector constrains the transition pattern. One aspect of this constraint is that transitions between features with high probability of viewing occur with corresponding high frequency (Senders, 1966; Senders, Grignetti, and Smallwood, 1966). Thus, a high frequency of transition itself is not a necessary indication of a statistical association of fixations on pairs of features (Carpenter and Just, 1978). A statistical technique described in the Appendix shows how to identify genuine statistical dependencies in transition patterns.

Somewhat surprisingly, oculomotor information seeking during a variety of tasks such as visual search (Engle, 1977; Inditsky and Bodmann, 1980; Kraiss and Knaeuper, 1983; Krendel and Wodinsky, 1960), instrument monitoring (Senders, 1966; Weir and Klein, 1970; Wewerinke, 1981), computer-menu scanning (Card, 1983), and during solution of seriation problems (Groner and Groner, 1982), may be modeled as random or stratified random sampling with replacement. This apparent randomness in visual scanning is especially surprising in view of evidence from other experiments, which shows that subsequent fixations may be directed by information acquired at the current fixation (Kapoula, 1983; Rayner and Pollatsek, 1981; Vaughn, 1982). If the underlying cognitive

[1] Requests for reprints should be sent to Stephen R. Ellis, NASA–Ames Research Center, MS 239-3, Moffett Field, CA 94035

processes directing eye movements during information seeking can be made periodic and statistically stationary, evidence of that direction ought to be evident as statistical dependencies in the observers' scanning during free viewing.

The following experiment was designed to exhibit these dependencies in an information-seeking task. Periodicity and stationarity of the processing of visual information in this task were encouraged respectively by periodicity in presentation and by extensive training of the subjects. We endeavored to measure the extent to which our subjects' scanning eye movements could be described as random, and thereby to infer the extent to which their control may be autonomous from repetitive ongoing cognitive processing. If the decision of where to fixate next were controlled by repetitive open-loop information-gathering strategies such as left-right scan as in reading (Bouma and de Voogd, 1974; Kolers, 1976), or by closed-loop strategies in which the decision is based on information gathered in the previous fixation (Rayner and Pollatsek, 1981), then scanning among points of interest should exhibit statistical dependencies and should not be truly random.

METHODS

In this experiment, we examine the spatio-temporal structure of scanning eye movements made by airline pilots viewing a cockpit display of traffic information or CDTI (previously reported by Palmer, Jago, and Dubord, 1981; also see Verstynen, 1980).

Display Conditions

A series of 24 track-up, moving map CDTI displays was generated on a calligraphic computer graphics system (Evans and Sutherland PS 1) previously described by Palmer and his associates (Palmer, Jago, Baty, and O'Conner, 1980; Palmer et al., 1981). Each display depicted an encounter between two aircraft, the pilot's own ship and an intruder, both flying at the same altitude. An example of the display is provided in Figure 1. The boldface labels did not appear on the display when viewed by the subjects. Each aircraft had 32 s of previously tracked positions displayed as eight dots of trail, one for each update, and had a 32-s predictor which indicated its future position if it did not maneuver. The miss distances for all encounters were set at 1846 meters (6000 feet) and the map position of ownship was updated every 0.1 s. The intruder's position was updated every 4 s. Equal numbers of intruders were randomly determined to pass in front of and behind ownship. Map range from ownship to the top of the display was set at 18.5 km (10 nautical miles). In addition to ownship and the intruder, the display contained two geographical locations, LOM and PEPSI, and a route shown as a solid line. All trajectories crossed near LOM. Each encounter consisted

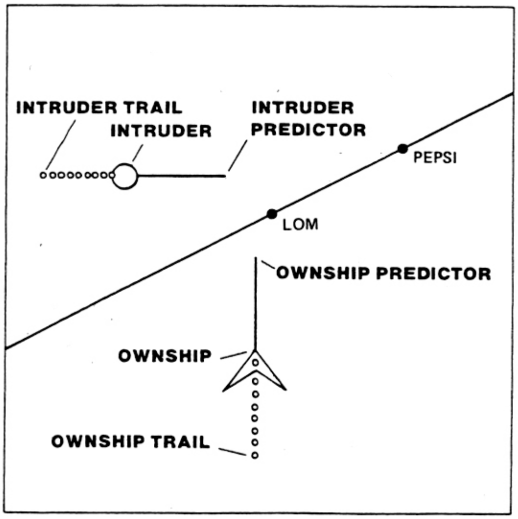

Figure 1. *Representative encounter between ownship and an intruder approaching from left.*

of seven 4-s updates simulating radar-derived data: two updates (8 s) before the intruder appeared, and five updates (20 s) afterwards. To avoid the mixing of different scan strategies, analysis was restricted to the final five updates during which the intruder was visible. The first two updates of the display did not provide sufficient data for a separate transition analysis of this phase of the experiment.

The display was blanked in such a way that, after the last update, there remained 44 s before the flight paths of the aircraft crossed. This blanking prevented the aircraft symbols from ever overlapping. After the display was blanked, the subject was prompted to decide if the intruder would pass in front of or behind ownship. The subject responded by pressing a two-way switch, was shown the correct response by text on the screen, and then proceeded to the next encounter. The particular encounters used provided a task of moderate difficulty as compared to those used in previous experiments (Palmer et al., 1980).

The encounters represented both straight and turning horizontal encounter geometries. Intruders approached randomly but equally often from the left and right of ownship. There were four different types of encounters, each producing different target movements on the display: neither aircraft turning, intruder only turning, ownship only turning, and both aircraft turning. Turn rate for all aircraft was constant at 1.5 deg/s. The order of presentation of the four types of target movement, 6 distinct encounters per type, was randomized within blocks of 24 encounters of the same predictor type. Each subject viewed two blocks, for a total of 48 distinct encounters. Significantly, the turning of ownship caused the other parts of the map to revolve around it.

The resulting set of encounters provided a wide variety of encounter geometries in terms of display coordinates. This variety was also intended to discourage stereotyped scanning, such as a left-to-right reading pattern, from masking a scanning strategy based on information present on the display. The variety was additionally intended to ensure that any scanning strategies were not consequences of the particular placement or movement of the aircraft and ground symbols on the display. Thus, any regularities identified in the scanning could be attributed to the information represented by the display elements and not to regularities in their placement on the display.

Videotapes of the encounters were made for off-line presentation. They were time-marked with signals on the audio channels, in order to establish synchrony between the records of eye movements taken while the subjects made in-front/behind judgments. The tapes were played back on a TV monitor so that the display subtended a rectangle of 12×10 deg with average luminance of about 1.0 cd/m^2. The outlines of the symbols had a luminance of about 3.0 cd/m^2. Complete aircraft symbols on the display subtended a visual angle of 3.5 deg. The monitor was viewed from a distance of 75 cm by the subject, who sat in the chair of a simplified part-task cockpit simulator. The chair had a high back and was fitted with a chin rest suspended from above to restrict head movement. Subjects signaled the in-front and behind judgments with a toggle switch. A separate push-button switch was used during calibration of the eye monitor to signal fixation of a calibration marker. Masking noise was provided by the sound of several motors associated with the eye tracker and PDP-12 computer used to record the data.

Direction of gaze data were recorded with a Gulf and Western 1994 pupilometer-based television eye monitor, which was calibrated by recording fixations at 25 reference points in a 5 × 5 array (14 deg/side), centered in the subject's forward field of view. The eye monitor performed within specification, providing at least 1 deg overall accuracy in measuring eye position. The eye monitor output (*x,y* direction of gaze and pupil diameter), the subjects' signals, and the time markers from the videotape were all digitally recorded. The sample rate was 30 hz, set by the video frame rate.

Subjects

Eight male airline pilots were subjects in the experiment. All were either captains or first officers with at least 6000 hours flying experience. All had at least three hours experience in similar CDTI experiments requiring in-front or behind judgments and had recorded better than average performance in these studies.

Procedure

During an orientation session before each experiment, the subject's attention was diverted from the fact that his direction of gaze was being recorded. He was told that the purpose of the experiment was to determine if pupillary changes could be used to predict projected-flight-path judgments. Lengthy briefing was unnecessary due to the subjects' experience in similar experiments. The meaning of all parts of the symbology was reviewed, however, and each subject was given about 20 minutes practice making in-front/behind judgments before his scanning patterns were recorded.

This training and the subject selection procedure resulted in asymptotic, near-perfect performance of the task, which provided stable behavior for analysis. After the initial practice, the eye monitor was adjusted to track the left eye, and an initial full calibration was made. Interspersed between data gathered during the encounters were reset fixations, taken by having the subject refixate a position corresponding to the center of the calibration grid. A reset calibration was taken when a drift of more than 1.0 deg was observed on the eye tracker's CRT display of eye position.

Data Processing

After the experiment, the data were transferred to a PDP-11/70 computer to be linearized according to a piecewise-linear approximation derived from the calibration fixations. Fixation locations were determined in a manner similar to that of Karsh and Breitenbach (1983), thereby identifying fixations representing at least 90 ms in duration.

After identifying the positions, durations, and onset times of all fixations, the data were correlated with records of the positions of all points of interest as a function of time after the beginning of each encounter. Thus, each fixation could be assigned to one of eight possible points of interest: the end of ownship's trail (OST), ownship present position (OS), the end of ownship's predictor (OSP), the end of intruder's trail (IT), intruder's current position (I), the end of intruder's predictor (IP), location PEPSI (PEP), and location LOM (LOM). All fixations not within 1 deg of any of these points of interest were assigned to a category called BIN. The data were then tabulated to determine the overall distribution of fixation durations as well as separate distributions for each point of interest. Percentage of time spent at each point of interest was determined.

RESULTS

Distribution of Fixations

The individual distributions of fixation durations had the positive skew—1.84 to 2.64—usually found in distributions of fixation duration. The means ranged through a region somewhat longer than that usually

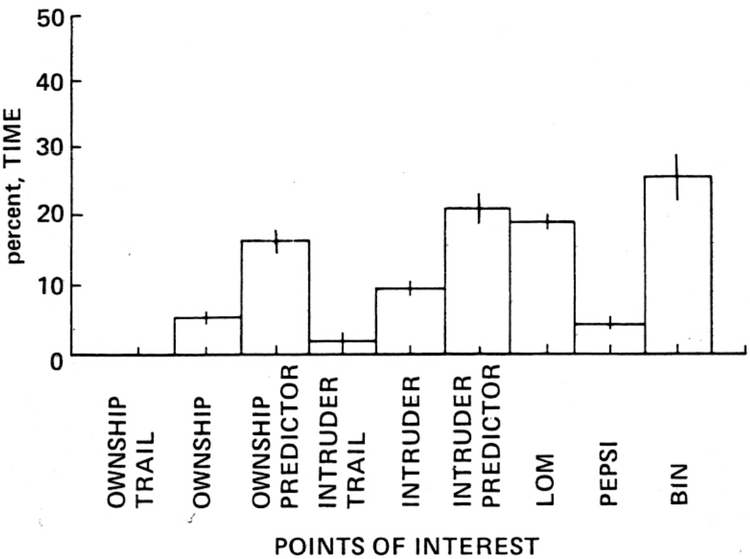

Figure 2. *Means across subjects of percentage of time spent viewing each point of interest (+/− 1 standard error).*

found for scanning of graphical stimuli—341 to 554 ms—and standard deviations ranged from 282 to 498 ms. All pilots distributed their fixations across the eight points of interest in a consistent manner, which in some respects supports earlier findings concerning the differential usefulness of the predictor lines and history dots attached to each symbol (Palmer et al., 1980). For example, the history dots displayed with all aircraft were almost never fixated. This observation is consistent with findings of previous experiments, in which the presence of these dots did not improve the accuracy of the front/behind judgments.

The higher proportion of viewing time on LOM, compared with PEPSI, probably occurred because LOM was close to the point of intersection of the flight paths for all the encounters. With the exception of LOM, the percentage of time each point of interest was viewed was approximately constant during the course of the encounter (within 8%). The time spent viewing LOM increased about 40% as a function of time after the appearance of the intruder; that is, as a function of each 4-s update period, $F(1,7) = 15.02$, $p < 0.01$. This increase in viewing of LOM corresponded to a decrease of about 30% in unclassified fixations, which were generally at points intermediate between the ends of the predictors on the aircraft.

Analysis of Transition Frequencies

We have examined the possibility that each subject's probability of viewing each point of interest can predict the frequency of transitions among them. To do so, we calculated expected frequencies of transitions (as described in the Appendix), and compared them with observed transition frequencies. This comparison was made on a subject-by-subject basis with a chi-square goodness-of-fit test on the entire distribution of observed and expected transitions. The method of analysis is illustrated for one subject in Table 1, which contains a matrix of his observed and expected first-order transition frequencies.

The observed transitions were determined from eye movement data collected while the subject viewed the 20-s periods of 48 different traffic encounters. Corresponding to each observed transition frequency, an expected transition frequency was calculated on the assumption of stratified random sampling (Appendix Equations 1 and 2). Provided that those cells with small expected frequencies—for example, $f_e(i \rightarrow j) < 5$— are collapsed, the observed and expected frequencies furnish a basis for calculating a chi-square goodness-of-fit test for the entire transition matrix. Several different methods of collapsing the cells with small expected frequencies were tried, and the results are insensitive to the choice of methods.

The underlined entries in Table 1 indicate which of the observed transitions were identified as sufficiently different from the expected values to be considered evidence for statistical dependency in the scanning. The main diagonal is undefined, since we are unable to observe a transition from a point of interest to itself. Minor irregularities in the pattern of transitions, such as exits from a point of interest that seems never to have been entered, are due to breaks in the sequence of eye movements when the eye tracker lost track of the eye.

The number of degrees of freedom for the goodness-of-fit test are $n(n-2) - 1$, where n is the number of points of interest. Two degrees of freedom are lost for each point because of exclusion of the main diagonal, and the fact that the number of visits to each point must equal the number of exits from it. In fact, because of breaks in recording the sequence of fixations, the number of visits to a particular point of interest usually did not exactly equal the number of exits from it.

As shown by the chi-square statistic at the bottom of Table 1, there is statistically significant deviation between the overall observed and expected transition patterns and, thus,

TABLE 1

A Single Subject's First-Order Transition Frequencies among 8 Points of Interest (Corresponding Expected Frequencies in Parentheses)

			To					
From	*OST*	*OS*	*OSP*	*IT*	*I*	*IP*	*LOM*	*PEP*
Ownship trail	—	1	0	0	0	0	0	0
(OST)	—	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Ownship	0	—	11	1	3	6	6	5
(OS)	(0)	—	(4)	(0)	(4)	(8)	(6)	(1)
Ownship predictor	0	12	—	0	8	14	30	5
(OSP)	(0)	(4)	—	(2)	(16)	(30)	(21)	(3)
Intruder trail	0	0	0	—	6	1	2	0
(IT)	(0)	(0)	(2)	—	(2)	(3)	(2)	(0)
Intruder	0	2	9	6	—	<u>47</u>	12	2
(I)	(0)	(4)	(16)	(2)	—	<u>(29)</u>	(23)	(3)
Intruder predictor	0	4	21	0	<u>54</u>	—	40	4
(IP)	(0)	(8)	(30)	(3)	<u>(30)</u>	—	(43)	(6)
Waypoint LOM	0	2	10	0	23	47	—	5
(LOM)	(0)	(6)	(24)	(2)	(23)	(43)	—	(5)
Waypoint PEPSI	0	5	7	1	3	7	7	—
(PEP)	(0)	(1)	(3)	(0)	(3)	(6)	(5)	—

Total number of transitions = 429
Chi-square = 178.8, df = 47, $p < 0.005$

there is evidence that something other than stratified random sampling is taking place during the scanning. Chi-square tests for seven of the eight subjects show highly reliable differences between the observed and expected transition frequencies (Table 2).

The chi-square values represent goodness-of-fit tests of stratified random sampling as a model of the scanning data. As shown by the significance levels in Column 3 of Table 2, this model can be rejected as a description of the scanning for seven of the eight subjects. However, as shown by the correlations between the observed frequencies of transition and those expected from stratified random sampling, the stratified random sampling model accounts for a good deal of the pattern of transitions. The correlations based on the log transformed data are also shown. The last two columns of Table 2 compare the statistical dependency of the observed scanning with that expected from stratified random sampling, and show that the observed scanning for all subjects is more statistically dependent than would be expected from stratified random sampling.

The one subject who did not show a reliable difference had a sparse transition matrix with the fewest transitions on which to base an estimate of the probability of a transition, p(i to j). His data were, in other respects, however, qualitatively similar to those of the other subjects.

For no subject, however, can the chi-square test alone address either the magnitude or the direction of the deviations from stratified random sampling. Accordingly, in order to assess the extent of the deviation, each subject's expected transition frequencies were regressed against his corresponding observed frequencies (Tukey, 1977). In such a regression (Figure 3), a perfect prediction corresponds to a linear regression with a slope of 1.0 and a correlation coefficient of 1.0. The slopes of the regressions of all subjects are quite close to the 1.0 (dashed line), and there is a strong linear relation. Figure 3 collapses this analysis across all subjects, resulting in the superposition of many points at the lower frequencies. Regressions were calculated separately for each subject and drawn through the scatter plot. Three of the regression lines are superimposed because of nearly identical parameters.

The strength of this relationship plotted in Figure 3 is measured by two correlations between observed and expected frequencies, shown for each subject (Table 2). The first, in Column 5, is the Pearson product-moment correlation, corresponding to the regression shown in Figure 3; that is, the correlation between the observed frequency and that ex-

TABLE 2

Results of Analyses of Each Subject's Entire First-Order Transition Matrix

Subjects	*Chi-square df = 47*	*Chi-square statistical significance*	*Number of transitions*	*Corr*	*Corr (df) log*	*H_c observed bits*	*H_c expected bits*
1	35.3	—	154	0.97	0.84 (39)	1.606	1.785
2	152.5	$p < 0.001$	417	0.96	0.79 (51)	1.940	1.980
3	134.8	$p < 0.001$	409	0.94	0.71 (43)	1.846	2.115
4	97.3	$p < 0.001$	348	0.95	0.75 (49)	1.838	1.978
5	82.1	$p < 0.005$	270	0.93	0.79 (43)	2.003	2.197
6	84.5	$p < 0.005$	431	0.97	0.84 (57)	2.282	2.509
7	178.8	$p < 0.001$	429	0.96	0.82 (48)	1.885	2.104
8	78.2	$p < 0.005$	275	0.94	0.72 (43)	2.002	2.183

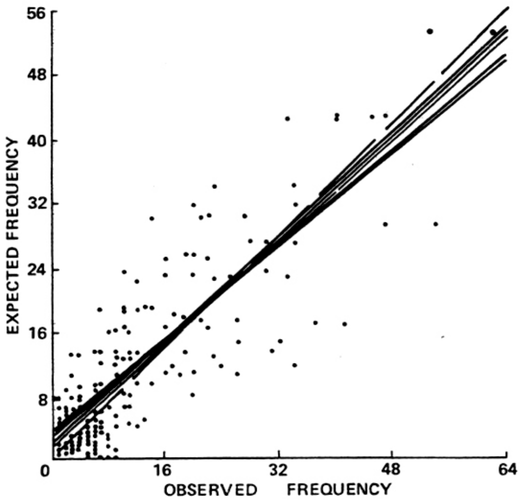

Figure 3. *Scatter plot of observed point-to-point transition frequencies versus those expected based on stratified random sampling.*

pected by assumption of random sampling. The second, in Column 6, is a Pearson correlation based on log transforms of both expected and observed frequencies, which corrects for the skew in the marginal distributions of observed and expected frequencies. As seen from Tables 1 and 2, though the stratified random sampling model used to calculate the expected frequencies provides an approximation of the empirical first-order transition pattern there are noticeable deviations from the expected values (Figure 3).

Identification of Statistical Dependencies

We assessed the direction of the actual deviations, on a subject-by-subject basis, by transforming their observed and expected first-order transition matrices of p(i to j) into a corresponding conditional probability matrix of p(i,j), and we calculated the conditional information, H_c (Appendix Equation 3), for each (Table 2). These calculations consistently indicated that the observed H_c of the transition matrices were lower in magnitude, and thus more statistically dependent, than the expected matrices (two-tailed sign test, $p < 0.008$). This contrast shows that the direction of the statistically significant, overall chi-square tests is toward more statistical dependency than that predicted by the stratified random sampling model.

To isolate terms in the chi-square calculations that contribute to the overall deviation, we treated each term as a separate test with one degree of freedom. For example, in Table 1, after the collapsing of the cells with small expected frequencies to a single cell, there remained 21 separate terms for the chi-square calculation. Thus, there are 21 separate tests, 20 of which may be treated as independent. If one adjusts the probability of the Type II error of omission for each test so that the entire analysis on each subject is kept within the customary 0.05, the separate terms may be tested for deviation from stratified random sampling.

The results of this procedure are shown in Table 1, in which underlined text is used to indicate the two transitions that reliably contribute to the overall statistical dependency in the scanning. Similar analyses were carried out on the transition patterns for the other seven subjects in the experiment (Figure 4). In essence, the analysis applies a filter to the transition patterns to identify transitions genuinely indicating statistically dependent "linkage" between information presented at a pair of points of interest. As shown in Figure 4, the transitions exhibiting statistically dependent associations do not necessarily correspond to those with the most frequent transitions. The left-hand panel of Figure 4 shows each subject's transition pattern among points of interest on the CDTI display. According to the legend, thickness of the arrows connecting pairs of points on these panels codes the relative frequency of each transition. Corresponding right-hand panels identify only those transitions that exhibit true statistical dependencies for each

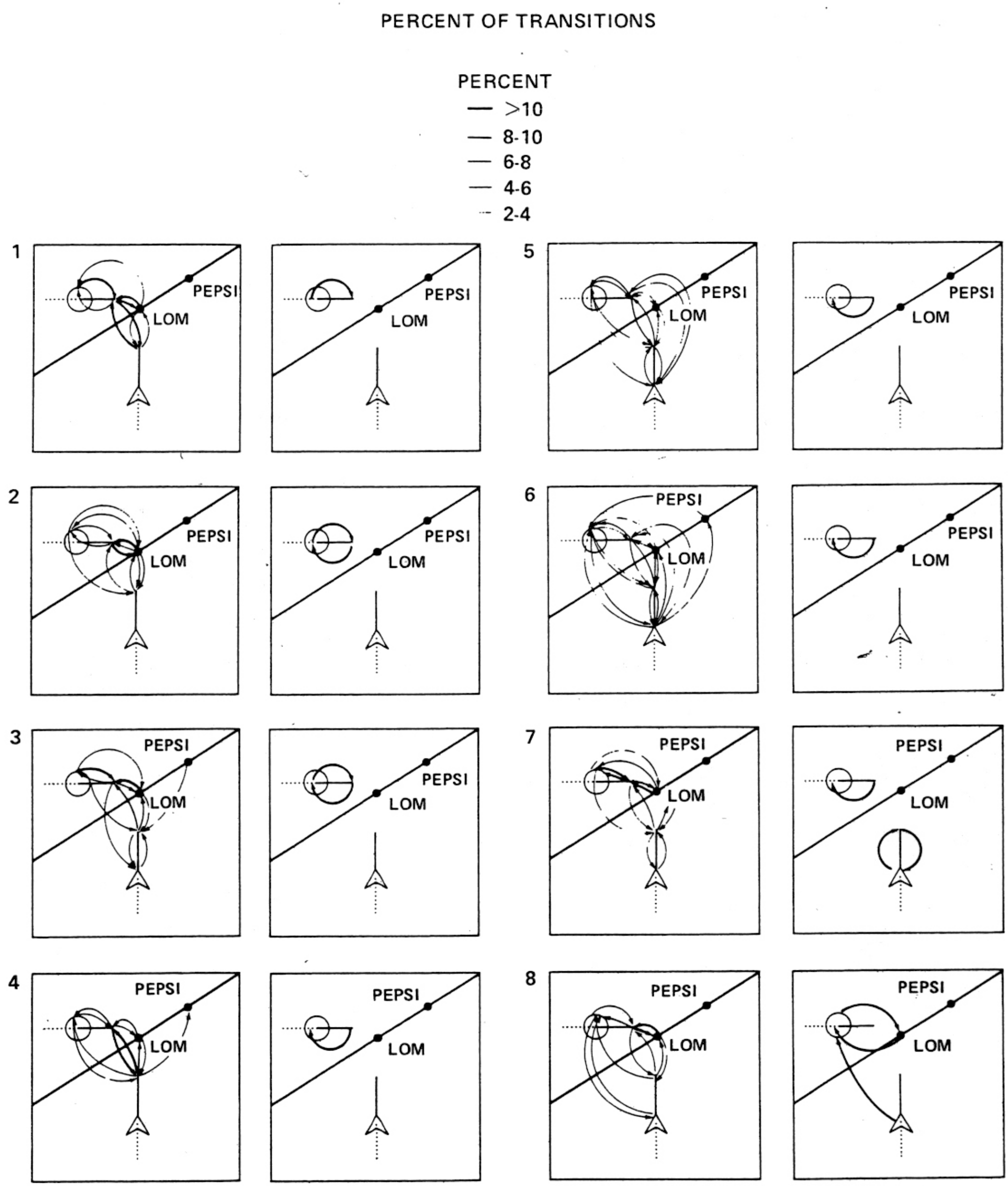

Figure 4. *Scanning patterns from each of 8 pilots who used the CDTI to estimate future relative position of two aircraft.*

respective subject. The thickness of the arrows in the right-hand panels has no special significance.

The particular transitions for which the observed transition frequencies deviated from those expected for stratified random sampling were similar across all pilots. This was shown by nonparametric Friedman ANOVA of the rank ordering of these deviations; Kendall W = 0.524, chi$(r)^2$ = 171.91, df = 41, $p < 0.001$. Most deviations were associated with transitions involving the aircraft predictors. The relatively small Kendall W reflects the fact that most of the agreement between subjects was confined to the few larger deviations in which the observed frequencies of transition exceeded the expected frequencies.

If the identified scanning dependencies were caused by closed-loop control, evidence for it could be found in the correlation between fixation duration and preceding saccade size. One view of this kind of control would be that information from a subsequent fixation position acquired during the current fixation influences the duration of that next fixation. If it were based on the greater peripheral preprocessing possible when the positions of subsequent points of interest fall within the functional field of view of the preceding fixations, this kind of interaction could produce a positive correlation between fixation duration and preceding saccade length. The closer a subsequent fixation is to the previous fixation, the more preprocessing would be possible. Accordingly, the fixation duration of the subsequent fixation may be reduced, since the information it can provide has already been partially processed (Kapoula, 1983).

We examined this aspect of our data subject by subject, and found no evidence for this kind of correlation. In fact, in our task there appeared to be a slight reverse effect. All of the subjects showed statistically significant negative correlations between previous saccade size and fixation duration. They ranged from -0.14 to -0.54 ($p < 0.05$ or better). Inspection of the scatter plots underlying these correlations and use of log transforms to correct for positive skew of fixation duration confirm these negative correlations. Breakdown of this analysis by update period, to check for changes during the course of an encounter, showed that the generally linear decrease of fixation as a function of preceding saccade size was present throughout the encounter. Figure 5 summarizes the decrease in mean fixation duration as a function of the increase in size of preceding saccade.

DISCUSSION

The original hypothesis in this experiment was that either repetitive open-loop or closed-loop information-gathering strategies would introduce statistical dependencies into the subjects' information-seeking scanning patterns. This hypothesis was confirmed. However, the results presented here also show that the deviation from a stratified random sampling model, though consistently in the direction of greater determinism and dependency, was small in magnitude (see

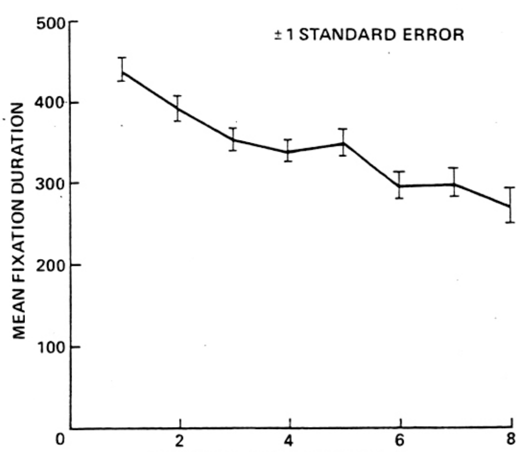

Figure 5. *Summary of the decrease in mean fixation duration as a function of the size of the preceding saccade.*

Table 2). What could account for the general randomness in the scanning?

Causes of Statistical Independence in Scanning

To the extent that the selection of the next fixation point is statistically independent of the previous fixation and describable as stratified random sampling, the results of this experiment are consistent with visual sampling models that assume that the search for information is autonomous of subsequent processes that use it (Didday and Arbib, 1972; Smallwood, 1967). These types of models represent the eye as a "buffer filler" that is not under the tight control of the higher-order information processing that underlies interpretation of visual information (Bouma and de Voogd, 1974; Kolers, 1976). Models like this could be likened to a real-time data acquisition system, such as DEC's RT11, which uses a direct memory access device to service an input buffer.

However, the collection of the eye movement data over time and across different encounter conditions raises the possibility that the apparent randomness of the data could arise as an artifact of the mixing of a variety of scanning strategies, making the transition pattern nonstationary. In a nonstationary process, genuine but changing statistical dependencies could be obscured by mixing in the overall analysis (W. F. Clement, personal communication, August 1981).

The evidence for such mixing is weak, since, with the exception of the percentage of time viewing LOM, the sampling probabilities remained relatively constant throughout the session. The change in the percentage of time viewing point LOM was probably an artifact created by LOM's being within 1 deg of visual angle of the intersection of the flight paths. This intersection was a point between the ends of the predictors and appeared to be a point fixated by the pilots during the encounter. As each encounter progressed, the pilots' fixations between the ends of the predictors, which were included in the BIN classification, moved closer to LOM. Thus, due to inevitable uncertainty in classification of fixations towards the end of each encounter, the percentage of viewing time assigned to LOM increased, and the percentage assigned to BIN decreased.

A more likely cause than the mixing of scan strategies for a stratified random sampling characteristic of the data would be the measures taken to ensure stationarity in the scanning. Such measures may have had the paradoxical effect of encouraging random sampling. In this view, the efforts to overtrain the pilots to asymptotic, near-perfect performance on the in-front/behind discrimination may have caused them to develop such efficient decision strategies that they were able to monitor several aspects of the encounter without consciously switching their attention.

Under these conditions, the display serves the pilot as a kind of random access memory, allowing immediate acquisition of any fact required for further analysis without the need for search (Groner and Groner, 1982). This is a situation describing the well-trained pilot, who is able simultaneously to monitor and control several different aircraft systems. Less well-trained pilots have to shift their attention consciously from one system to another, from one display to another, and may exhibit considerably more statistical dependency in their scanning eye movements (DeMaio, Parkinson, Leshowitz, Crosby, and Thorpe, 1976). Indeed, as has been suggested before (Ellis and Stark, 1981; Tole, Stephens, Vivaudou, Harris, and Ephrath, 1982), anything that interferes with a pilot's ability to monitor different dynamic systems, such as stress or workload, might increase the statistical dependency in his scanning eye movements. Thus, measures of statistical dependencies in scanning eye movements may provide useful indices of workload or stress. The conclusion to be drawn is that statistical dependencies, whatever their practical uses, are detectable and need explanation.

Reasons for Statistical Dependencies in Scanning

In general, two causes of statistical dependencies may be distinguished: closed-loop control and open-loop control. In closed-loop control, information acquired during fixation is used to direct the subsequent saccade. In open-loop control, direction of the next saccade is due to information processing independent of the current visual information in the visual field.

In visual tasks that explicitly benefit from peripheral preprocessing, evidence has been presented, for example, by Kapoula (1983), to show that fixation durations on subsequent points of interest can be influenced by their proximity to previous fixations. This kind of evidence clearly suggests that sequences of fixations are influenced by closed-loop control processes. This result is also in accord with classical observations that the peak velocity of a saccade is proportional to saccadic amplitude (Bahill and Stark, 1979), the so-called main sequence law. In that case, the controller driving the eye must "know" the coordinates of the end of the saccade before it generates the move commands to the muscles.

Our fixation data, however, do not exhibit the same relationship between fixation duration and previous saccade size that was offered by Kapoula as evidence for the directing role of peripheral preprocessing, a role that illustrates closed-loop control. Our subjects' task, however, was quite different from that used by Kapoula, and this difference may explain the different results. In her experiment, the subjects made a fine visual discrimination requiring high-resolution foveal vision. In this task, peripheral preprocessing of the target of the next fixation could be reasonably expected to reduce the next fixation's duration, since some aspects of the target's identity would be known before the next fixation is made.

In our experiment, subjects were required to judge relative positions of two targets and would benefit from simultaneous viewing of both. Accordingly, the longer fixation durations we found to be associated with the shorter preceding saccades could reflect this simultaneous processing of both targets' positions. When the targets are close together and both can be included in the functional field of view, fixations are longer, allowing processing of the positions of both targets. The fixations to the more widely separated parts of the displays would not necessarily allow simultaneous processings of the current and previously viewed target's positions. Thus, these fixations would be of shorter duration compared with those preceded by shorter saccades. Accordingly, despite the differences from Kapoula's results, our results are consistent with a model in which peripheral information influences fixation duration, thereby implying closed-loop control.

An alternative cause of statistical dependency could be called open-loop control. An example might be the left-to-right scanning used in reading English text. This asymmetry is not a consequence of the specific information present in the text (though the specific fixations may be), but only of the habitual lexicographic layout. Such open-loop scanning is not necessarily under the control of ongoing information processes; it may be simply filling an input buffer. However, open-loop control of scanning need not be based on scanning habits. Internal information processing, in which the processes themselves drive the subjects' sequence of fixations, could control the scanning. An example of this might be movements of the eye to answer "visual questions" raised by previous viewing (Hochberg and Brooks, 1978), or scanpaths hypothetically linked to the memory trace of an object (Noton and Stark, 1971) or to cognitive models (Ellis and Stark, 1978, 1979).

To differentiate between open- and close-loop control, one would have to determine the effect of the spatial location of the points of interest on statistical dependencies identified in the scanning. If it could be shown that the closer points of interest in our experiment were more likely to be the end points of statistically dependent transitions, one could then argue that the dependencies were due to peripheral preprocessing that had taken place during the preceding fixation. If involvement of a point of interest in a dependent transition were not related to its proximity to another point, the dependency could then be attributed to open-loop control. The data from the present experiment do not provide sufficient numbers of true dependencies for this analysis.

The principal conclusion to be drawn from this experiment is that random sampling models of visual information seeking in dynamic visual environments, though very good approximations, do not completely account for the pattern of information seeking we observed. As emphasized by the Friedman ANOVA, which shows consistent deviations from stratified random sampling, a model incorporating some determinism is required.

Furthermore, the above analysis underscores an important caveat for those wishing to interpret scanning patterns: the probability of sampling points of interest in a pattern must first be considered before transitions of fixations among them can be interpreted as evidence of "linkage" of the information present at the points (Senders, 1966). Variation in these sampling probabilities can result in scanning patterns that, because of frequent transitions between popular points of interest, appear to exhibit statistical dependencies, but in fact do not. A technique like the one described in this paper is necessary to identify those instances of true statistical dependency actually in the eye movement data.

APPENDIX

To understand what is meant by statistical dependency in visual scanning, it is helpful to consider three alternative modes of scanning among points of interest (Figure 6).

The random case in Figure 6 illustrates completely unconstrained scanning among three points of interest. The stratified random case shows the effects of scanning when the sole constraints are provided by the differential probabilities of viewing each point of interest, p(i) in the text. In this case, the observed pattern of transition is exactly that which could be calculated from the p(i) by Equation 1. The statistically dependent case, which shares the same p(i) as the stratified random case, illustrates how the observed pattern of transitions may deviate from that calculable from the p(i). Note that the larger deviations between the observed transitions and those consistent with stratified random sampling are not necessarily associated with the most frequent transitions.

The description of these different modes of scanning first requires the distinction of three types of probabilities:

(1) p(i), p(j), or p(l), the simple probabilities of viewing points of interest i, j, or l respectively. Although this probability may be defined by the number of times a point of interest is visited, in this paper it is defined by the percentage of time spent on each particular point. The conclusions of this paper, however, are insensitive to the choice of definition.

(2) p(i to j), the probability of a transition between distinct points of interest i and j, con-

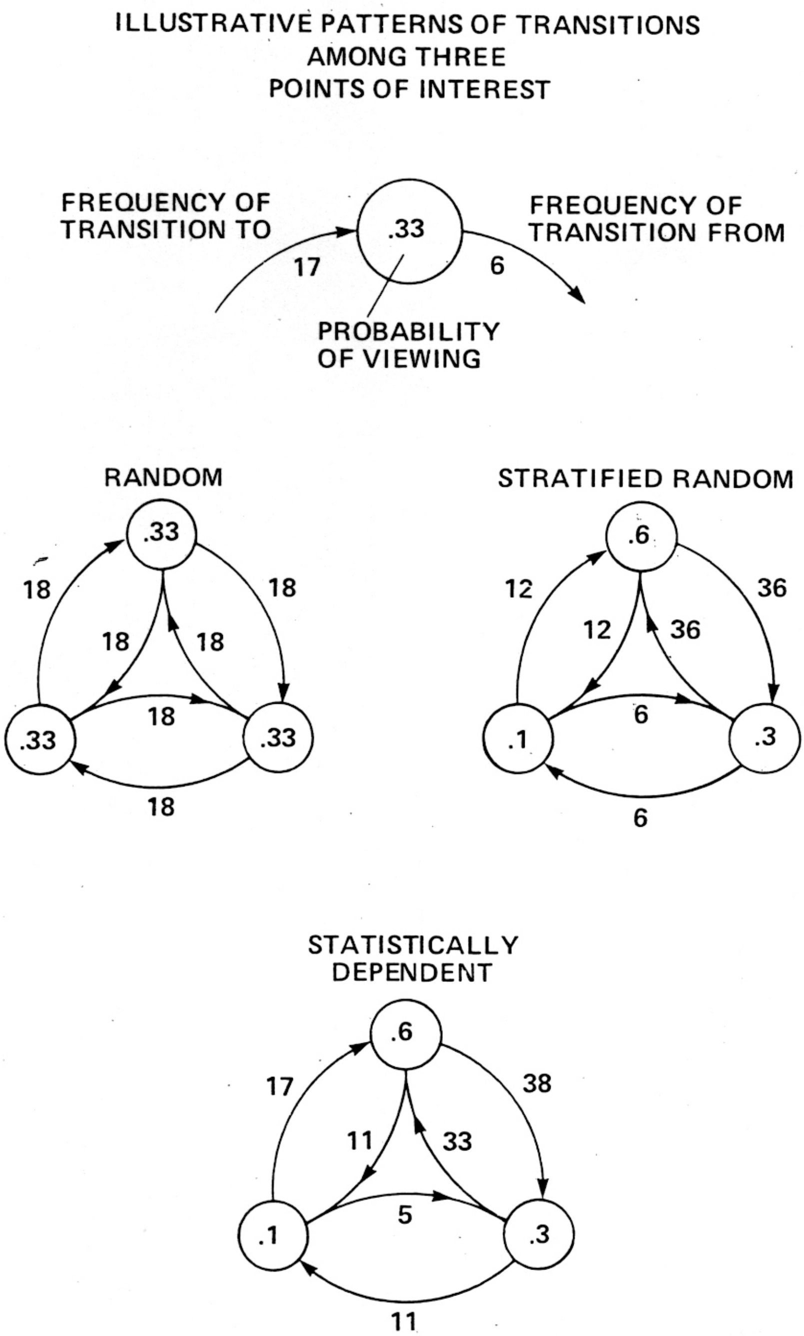

Figure 6. *Examples of different types of scanning among three points of interest.*

ditioned only on the assumption that unobserved transitions occur from each point to itself (i.e., the link value)

(3) p(i,j), the conditional probability of viewing point of interest j given previous viewing of point of interest i.

The latter two probabilities, p(i to j) and p(i,j), may be related to each other by p(i,j) = [N/n(i)] p(i to j), where N is the total number of transitions among all points of interest and n(i) is the number of exits from a particular point i.

These three probabilities allow the description of the three modes of scanning among points of interest in order to sample visual information. Visual sampling of points of interest may be completely random, stratified random, or statistically dependent. All sampling is assumed to be done with replacement. In the random case, each point is viewed with equal probability, and, as a consequence, all transition probabilities between pairs of points are equal. In the stratified random case, the points of interest may be viewed with different probabilities. This distribution may be dependent on the particular task the viewer must undertake, but now only the transitions to and from particular pairs of points need be equal.

Note that, by chance alone, there are many transitions between the more probable fixation points of interest. This case may be described alternatively as a zero-order Markov process, and thus the probability of fixating any point of interest is statistically independent of fixation on the preceding point. Accordingly, p(i to j) or p(i,j) are calculable from p(i) and p(j) (Senders, 1966). Only in a statistically dependent case (bottom Figure 6) do some of the transition frequencies illustrate the deviations from statistical independence that indicate statistical dependencies characteristic of a first order or higher Markov process. In this case, p(i to j) or P(i,j) are not calculable from p(i). It is also noteworthy, in this example, that the illustrated dependencies are not necessarily the most frequent transitions.

In the random case, sampling among the points of interest is completely unconstrained. In the stratified random sampling case, however, a differential probability of viewing the various points of interest leads to a constraint on the scanning sequences, which may produce the impression of sequential scanning. Under these conditions, transitions between high probability points of interest are likely due simply to the zero-order probability of viewing the respective points. Accordingly, any claim for statistical dependency in the transition patterns among points of interest must first show that the extent of the periodicity exceeds that which would be produced by the zero-order probabilities.

In order to detect truly statistically dependent transitions in scanning eye movement data, we have adapted an equation cited by Senders et al. (1966) for describing stratified random sampling. They noted that the joint assumption of (1) statistical independence of the transitions and (2) the existence of unobserved transitions from each point of interest to itself provides a means of calculating p(i to j)—the probability of a transition between any two distinct points of interest, i,j—provided that p(i), p(j), the zero-order probabilities of the two points are known.

$$p_e(i \text{ to } j) = \frac{p(i)p(j)}{1 - \sum_{l=1}^{n} p(l)^2}, \; i = j \qquad (1)$$

The denominator of this expression corresponds to the probability of all observable transitions between distinct points of interest. (Senders et al., 1966). Most importantly, Equation 1 provides a way to calculate expected transition frequencies, f_e (i to j), between points of interest, based solely on the expected transition probabilities and the total number of observed transitions, N.

$$f_e \text{ (i to j)} = Np_e \text{ (i to j)} \tag{2}$$

These calculated expected frequencies may then be compared by chi-square tests with the observed transition frequencies, in order to assess the adequacy of stratified random sampling as a description of the data. The number of degrees of freedom in these chi-square tests is $n(n-2) - 1$, if n is the number of points of interest. Two degrees of freedom were lost at each point of interest, because we were unable to observe transitions from a point to itself, and the nature of transitions requires that the number of visits to each point of interest equals the number of exits. This latter property was only approximately true of the data because of interruptions in recording of eye position when the eye monitor lost track of the eye.

A chi-square test comparing observed and expected transition frequencies that indicates a deviation is, however, a nondirectional test. The observed distribution could either be more uniform than that expected, so that the transitions in any row of the first-order matrix are more equal to each other than predicted by stratified sampling, or be less uniform than expected. The former deviation (more uniform) would be in the direction of less than expected statistical dependency. Conversely, if the observed distribution in any row of the transition matrix was less uniform, with larger maxima and smaller minima, such a deviation would be in the direction of more statistical dependency. The most statistically dependent case would be that with only one type of transition occurring from each point of interest. Consequently, differentiating between these two types of deviations requires a measure of the amount of statistical dependency in a transition matrix. Such a measure would allow comparison of the amount of statistical dependency in the observed transition matrix with that in the transition matrix expected from stratified random sampling.

The amount of statistical dependency in any transition matrix of p(i to j) may be measured by transforming it into a conditional probability matrix of p(i,j). The total conditional "information" (Brillouin, 1962) in the matrix, H_c, can then be calculated.

$$H_c = -\sum_{i=1}^{n} p(i)\left[\sum_{j=1}^{n} p(i,j)\log_2 p(i,j)\right], \; i \neq j \tag{3}$$

This value provides a measure of statistical dependency in the spatial pattern of fixations represented by the transition matrix, and may be used to compare one matrix with another. It has a maximum when the transitions from each point of interest are equally distributed to all other points. It has a minimum when the transitions from each point all uniquely go to a single point. Thus, the larger the H_c, the less the statistical dependency in scanning (Huff, 1966). For purposes of comparing the uncertainty in the spatial pattern of scanning, this measure is probably better than an alternative proposed by Tole et al. (1982), which explicitly combines fixation rate with uncertainty of scanning, so that the resulting "entropy of the scanpath" does not uniquely reflect the uncertainty in the spatial pattern of fixations.

This measure of statistical dependency is, however, significantly affected by the zero-order probabilities. For example, if only a few points of interest dominate the fixation distribution, there may be a small H_c but no genuine statistical dependency. Accordingly, like other statistics, sampling distributions for it must be determined so that its use may be extended to comparisons of distributions with different zero-order probabilities.

ACKNOWLEDGMENTS

Preliminary reports of the research contained in this paper were presented at the 17th Annual Conference on Manual Control, June 16-18, 1981 (JPL Publication 81-95), and at the 1982 Meeting of the Human Factors Society. The authors are pleased to acknowledge programming assistance provided by James A. Woods, Robert Krones, and Stephen Gonick of Informatics, Inc., and assistance in collection of data provided by Edward Denz and Bruce Hornstein. The research of Lawrence Stark was partially supported by the NASA cooperative agreement NCC 2-86 from the NASA–Ames Research Center.

REFERENCES

Bahill, A., and Stark, L. (1979, January). Trajectories of saccadic eye movements. *Scientific American, 240*, 84-93.

Bouma, H., and de Voogd, A. H. (1974). On the control of eye saccades in reading. *Vision Research, 14*, 274-284.

Brillouin, L. (1962). *Science and information theory* (2nd ed.). New York: Academic.

Buswell, G. T. (1935). *How people look at pictures*. Chicago: University of Chicago Press.

Card, S. (1983). Visual search of computer command menus. In H. Bouma and D. Bouwhuis (Eds.), *Attention and performance X* (pp. 97-108). Hillsdale, NJ: Erlbaum.

Carpenter, P., and Just, M. (1978). Eye fixations during mental rotation. In J. W. Senders, D. Fisher, and R. A. Monty (Eds.), *Eye movements and higher psychological functions* (pp. 115-134). Hillsdale, NJ: Erlbaum.

DeMaio, J., Parkinson, S., Leshowitz, B., Crosby, J., and Thorpe, J. A. (1976, June). *Visual scanning: comparisons between student and instructor pilots* (AFHRL-TR-76-10, AD-A023634). Williams Air Force Base, AZ: Air Force Human Resources Laboratory.

Didday, R. L., and Arbib, M. A. (1972). *Eye movements and visual perception: A 2 visual system model.* (COINS Technical Report 73C-9) Amherst, MA: University of Massachussetts.

Ellis, S. R., and Stark, L. (1978). Eye movements during the viewing of necker cubes. *Perception, 7*, 575-581.

Ellis, S. R., and Stark, L. (1979). Reply to Piggins. *Perception, 8*, 721-722.

Ellis, S. R., and Stark, L. (1981, June). Pilot scanning patterns while viewing cockpit displays of traffic information. In *Proceedings of the 17th Annual Conference on Manual Control* (pp. 517-524) (JPL Publication 81-95). Pasadena, CA: Jet Propulsion Laboratory.

Engle, F. L. (1977). Visual conspicuity, visual search, and fixation tendencies of the eye. *Vision Research, 17*, 95-108.

Fisher, D. F., Monty, R. A., and Senders, J. W. (Eds.). (1981). *Eye movements: cognition and visual perception.* Hillsdale, NJ: Erlbaum.

Fitts, P. M., Jones, R. E., and Milton, J. L. (1950). Eye movements of aircraft pilots during instrument landing approaches. *Aeronautical Engineering Review, 9*, 1–16.

Groner, R., and Groner, M. (1982). Towards a hypothetico-deductive theory of cognitive activity. In R. Groner and P. Fraisse (Eds.), *Cognition and eye movements* (pp. 181-195). Amsterdam: North-Holland.

Hochberg, J., and Brooks, V. (1978). Film editing and visual momentum. In D. Fisher, R. A. Monty, and J. W. Senders (Eds.), *Eye movements and the higher psychological processes* (pp. 293-316). Hillsdale, NJ: Erlbaum.

Huff, E. M. (1966). *Markov analysis of response timing on a drl schedule.* Unpublished doctoral dissertation, Texas Christian University, Ft. Worth, TX.

Inditsky, B., and Bodmann, H. W. (1980). Quantitative models of visual search. In *Proceedings of the 19th Symposium of CIE* (pp. 197-201). Paris: Commission Internationale de l'Eclairage.

Karsh, R., and Breitenbach, F. W. (1983). Looking at looking: The amorphous fixation measure. In R. Groner, C. Menz, D. Fisher, and R. A. Monty (Eds.), *Eye movements and psychological processes: International views* (pp. 53-64). Hillsdale, NJ: Erlbaum.

Kapoula, Z. (1983). The influence of peripheral preprocessing on oculomotor programming. In R. Groner, C. Menz, D. Fisher, and R. A. Monty (Eds.), *Eye movements and psychological functions: International views* (pp. 101-114). Hillsdale, NJ: Erlbaum.

Kolers, P. A. (1976). Buswell's discoveries. In R. A. Monty and J. W. Senders (Eds). *Eye movements and psychological processes* (pp. 373-396). Hillsdale, NJ: Erlbaum.

Kraiss, K. F., and Knaeuper, A. (1983). Using visual lobe area to predict visual search time. *Human Factors, 24*, 673-682.

Krendel, E. S., and Wodinsky, J. (1960). Search in an unstructured visual field. *Journal of the Optical Society of America, 50*, 562-568.

Noton, D., and Stark, L. (1971). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research, 11*, 929-942.

Papin, J. P., Naureils, P., and Santucci, G. (1980, May). Pickup of visual information by the pilots during a ground control approach in a fighter aircraft simulator. *Aviation, Space, and Environmental Medicine*, 463-469.

Palmer, E. A., Jago, S., Baty, D. L., and O'Connor, S. (1980). Perception of horizontal aircraft separation on a cockpit display of traffic information. *Human Factors, 22*, 605-620.

Palmer, E. A., Jago, S. J. and DuBord, M. (1981, June). Horizontal conflict resolution maneuvers with a cockpit display of traffic information. In *Proceedings of the 17th Annual Conference on Manual Control* (pp. 51-62). Pasadena, CA: Jet Propulsion Laboratory.

Rayner, K., and Pollatsek, A. (1981). Eye movement control during reading: Evidence for direct control. *Quarterly Journal of Experimental Psychology, 33A*, 351-373.

Senders, J. W. (1966). A re-analysis of the pilot eye-movement data. *IEEE Transactions on Human Factors in Electronics, HFE 7*, 103-106.

Senders, J. W., Fisher, D. F., and Monty, R. A. (Eds). (1978). *Eye movements and the higher psychological processes.* Hillsdale, NJ: Erlbaum.

Senders, J. W., Grignetti, M. C., and Smallwood, R. (1966, January). *An investigation of the visual sampling behavior of human observers* (NASA Contractor Report 434). Langley, VA: NASA Langley Research Center.

Smallwood, R. D. (1967, September). Internal models and the human instrument monitor. *IEEE Transactions on Human Factors in Electronics. HFE-8*, (3), 181-187.

Stark, L., and Ellis, S. R. (1981). Scanpaths revisited. In D. F. Fisher, R. A. Monty, and J. W. Senders (Eds.), *Eye movements: Cognition and visual perception* (pp. 193-226). Hillsdale, NJ: Erlbaum.

Tole, J., Stephens, A. T., Vivaudou, M., Harris, R. L., Sr., and Ephrath, A. (1982). Entropy, instrument scan, and pilot workload. In *Proceedings of the International Conference on Cybernetics and Society* (pp. 588-592). New York: IEEE.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Vaughn, J. (1982). Control of fixation duration in visual search and memory search. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 709-723.

Verstynen, H. A. (1980, May). Potential roles for the cockpit traffic display in the evolving ATC system. In *Proceedings of the International Air Transportation Meeting* (SAE Technical Paper #800736). Warrendale, PA: Society of Automotive Engineers.

Weir, P. H., and Klein, R. H. (1970, June). *Measurement and analysis of pilot scanning and control behavior during simulated instrument approaches* (NASA Contractors Report 1535). Moffett Field, CA: NASA–Ames Research Center.

Wewerinke, P. H. (1981). A model of the human observer and decision maker. In *Proceedings of the 17th Annual Conference on Manual Control* (pp. 557-570). (JPL Publication 81-95), Pasadena, CA: Jet Propulsion Laboratory.

Yarbus, A. (1967). *Eye movements and vision.* New York: Plenum.